DOCUMENT RESUME

ED 457 714 HE 034 360

AUTHOR Chen, Chau-Kuang; Campbell, Vickie C.; Suleiman, Ahmad
TITLE Predicting Student Performances at a Minority Professional School. AIR 2001 Annual Forum Paper.
PUB DATE 2001-00-00
NOTE 26p.; Paper presented at the Annual Meeting of the Association for Institutional Research (41st, Long Beach, CA, June 3-6, 2001).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Higher Education; Medical Education; *Medical Students; *Minority Groups; Models; *Prediction; *Professional Education; Program Effectiveness
IDENTIFIERS United States Medical Licensing Examination

ABSTRACT

This study attempted to build the best fitting prediction model, with high predictive validity, for success at a minority professional school. The resulting model could be used to document the effectiveness of academic programs and applicant screening. Data for 216 medical students from the college were used to evaluate the usefulness of the United States Medical Licensure Examination (USMLE) Step 1 pass status and test scores for this purpose. The USMLE models were constructed through the application of logistic and linear regression models. These models appeared reasonable and workable for the college because the significant predictors, the Medical College Admission Test (MCAT) scores, medical school freshman grade point average, sophomore course performance, and financial aid work-study dollars, were identified and included in the prediction equation. The measure of model goodness of fit and the overall prediction accuracy of the USMLE mode were reasonably high, and the assessment of the underlying assumptions of linear regression showed that linearity, normality, and independence were not violated. Ranking the predicted USMLE Step 1 scores and pass/fail status for prospective test takers, the administrators of the medical school could identify a small group of potential at-risk students and enable them to participate in mandatory board review or tutorial programs. Medical students themselves could use the prediction results to determine the optimum time to take the licensure examinations. (Contains 4 tables and 11 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 457 714

PREDICTING STUDENT PERFORMANCES AT A MINORITY PROFESSIONAL SCHOOL

By

Chau-Kuang Chen, Ed.D., Vickie C. Campbell, M.S., and Ahmad Suleiman, M.S.

Address:
Chau-Kuang Chen
Meharry Medical College
Office of Institutional Research
1005 DB Todd Blvd
Nashville, TN 37208
Email: CKCHEN@MMC.EDU
Phone: (615)327-6848
Fax No. (615)327-6328

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IR 031 360

BEST COPY AVAILABLE

PREDICTING STUDENT PERFORMANCES AT A MINORITY PROFESSIONAL SCHOOL

Chau-Kuang Chen, Ed.D., Director of Institutional Research, Meharry Medical College

Vickie C. Campbell, M.S., Director of Teaching and Learning Resources Center, Meharry Medical College

Ahmad Suleiman, M.S., Programmer Analyst, Meharry Medical College

PREDICTING STUDENT PERFORMANCES AT A MINORITY PROFESSIONAL SCHOOL

Abstract

The prediction models for the United States Medical Licensure Examination (USMLE) Step 1 pass status and test score with fifteen candidate explanatory were constructed by the application of logistic and linear regression methods. These models appeared to be reasonable and workable because the significant predictors--the Medical College Admission Test (MCAT) scores, medical school freshman GPA, sophomore course performance, and financial aid work-study dollar--were identified and included in the prediction equations. Also, the measure of model goodness of fit, namely R square value of .67 and the overall prediction accuracy of 80% were reasonably high. In addition, the assessment of the underlying assumptions of linear regression showed that linearity, normality, and independence were not violated.

Effective basic sciences education and financial aid support programs could be documented by the College's Institutional Effectiveness Committee since medical school freshman GPA, sophomore course performance, and work-study dollar contributed to the Step 1 performance. Additionally, the Admissions Committee could screen the qualified student applicants for interviews based on the additional knowledge of the relative influence (slopes or odds) of MCAT scores on the Step 1 performance. Ranking the predicted USMLE Step 1 scores and pass/fail status for prospective test takers, respectively, the administrators of medical school could identify a small group of potential at-risk students to participate in the mandatory board review or tutorial programs. Furthermore,

some medical students could use the prediction results to make the determination of when the optimum time would be to take the licensure examination. The prediction models could provide information to help the college enhance its effective academic and support programs and increase the likelihood of student success.

Introduction

Since the early 1990s medical students in the United States have been required to pass the USMLE Step 1 for progression to sophomore or junior levels in pursuit of a clinical sciences education. The USMLE Step 1 performance provided useful information regarding the knowledge and skills possessed by medical students, and when properly used it, was an important indicator of the quality and relevance of instruction received by these students (O'Donnell, 1993). The Step 1 is a standardized test that measures basic sciences knowledge and intends to require increased levels of students' critical thinking skills while reducing emphasis on recall of information (Erdmann, 1993 and Swanson, 1996). It emphasizes problem-solving skills in basic science disciplines essential to clinical medicine. Therefore, the USMLE Step 1 has become an important standard outcome measurement for effective medical education.

Passing the USMLE Step 1 is an important step in the medical licensing process, thereafter medical students are eligible for taking subsequent examinations Step 2 and 3. The Step 1 test score is widely used as a criterion for estimating the validity of the Medical College Admission Test (MCAT) and undergraduate grade point average (GPA) that are traditionally used to screen medical

school applicants for an admission interview (Elam, 1994 and Silver and Hodgson, 1997). Because of the significant value for improving medical education program and admission processes, there have been numerous studies investigating predictors of student performance on the USMLE Step 1 and utilizing modeling techniques to build the prediction models for licensure examination.

Among the variables being investigated as influencing factors on the USMLE Step 1, student performance in the first two years of medical school is considered the most prominent. The pre-admission variables such as undergraduate GPA and MCAT scores are also the two commonly used factors for building the prediction models. Among the statistical techniques being applied, least-square regression is the most popular method used to select the significant variables contributing to student performance on the Step 1.

The vast majorities of research studies are able to construct and interpret the functional relationship between various predictors and student performance on Step 1 successfully. However, some prediction models built for particular institution levels have flaws. For example, they include just a few independent variables in the models resulting in less explanatory power to describe the functional relationship between predictors and the Step 1 outcome variable. Having fewer independent variables in the models also lead to the low predictive validity. In several instances, researchers only use simple linear regression rather than the powerful multiple linear regression, which can simultaneously explain the relationship between multiple predictors and the USMLE Step 1 performance. In most cases, researchers only use a single technique to build their prediction models resulting in the inability to cross validate their model structure. Furthermore, many studies

fail to report the prediction accuracy and examine if the model assumptions of linearity, normality, and independence are violated.

The major emphasis of this study was to build the best fitting prediction model along with the higher predictive validity for a minority professional school. The resulting model could be used to document the effectiveness of academic programs and applicant screening processes.

Both logistic and linear regression methods were utilized in constructing the prediction models for medical licensure examination performances. Logistic regression was adopted because the outcome variables of interest consisted of dichotomous results, either passing or failing licensure examination. In other words, logistic regression analysis was involved to determine the probable influence of fifteen independent variables on the likelihood of passing USMLE Step 1. The objective of the maximum likelihood estimation is to find better approximations of the logistic regression coefficients that satisfy the maximum likelihood equation. Linear regression analysis was used to examine the functional relationships between USMLE Step 1 score and fifteen independent variables. The dependent variable, USMLE Step 1 score, is continuous on the measurement scale. The objective of least-square estimation is to find the regression coefficients that minimize the sum of squared distance from the observed to the predicted values of the dependent variable.

In this study, the prediction models are used to address the following research question frequently asked by faculty and administrators: "How well can the USMLE Step 1 pass/fail status or test scores be predicted by independent variables such as gender, ethnicity, the Historical Black

Colleges and Universities (HBCU) status, curriculum track (4- or 5-year track), undergraduate basic sciences average (BSA), undergraduate grade point average, the Medical College Admission Test scores, medical school freshman GPA, numbers of courses failed in the second-year curriculum, and amounts of student financial aid received?"

Literature Review

Predicting academic performance is a challenging task that requires the knowledge of modeling techniques and the availability of measurable predictor and response variables. A number of research papers published in recent years focused on identifying explanatory variables for student performances along with presenting the model goodness of fit as measured by the coefficient of determination (R square). The R square is also a measure of success of the regression model in explaining the variation in the data. It can be interpreted as the percent of the variation in student performance explained by all independent variables in the model. Therefore, a larger R square is desired in building reasonable and workable prediction models.

The MCAT scores and premedical GPA, when used together were considered to be important predictors of cognitive ability with R square value .42 for medical students (Shen and Comrey, 1997). Linear regression method was applied to derive the MCAT's predictive validity for student performances during the first two years of medical school. The study found that MCAT scores appeared to have slightly higher correlation (median R squares ranging from .38 to .44) with medical school grades than undergraduate GPA (median R square ranging from .29 to .33) (Koenig

and Wiley, 1996). Clearly, the research studies in the literature demonstrated that MCAT scores and undergraduate GPA were correlated with medical school performances in a positive direction.

The extent to which MCAT scores predict USMLE Step 1 performance was examined. The major finding indicated the MCAT is much more strongly related to USMLE Step 1 (R square .52) than undergraduate GPA (R square .23) (Koenig and Wiley, 1996). The significant Pearson correlation between MCAT and USMLE Step 1 scores across student subgroups including majority men (R square .25), majority women (R square .12) and minority women (R square .40) (Fadem, et al. 1995). MCAT scores among 112 medical schools provided more accurate predictions of the USMLE Step 1 performance (R square .32) than undergraduate information alone (R square .18). MCAT scores should continue to have substantial utility in the admission process, particularly in screening applicants to be interviewed (Swanson, et al. 1996). Furthermore, on average, each one point increase in average MCAT score resulted in a 7.62 point increase in USMLE Step 1 score with students from medical schools that require passing Step 1 for promotion performing slightly better. The largest correlation was for biological sciences (R square .29), followed by physical sciences (R square .26) (Swanson, et al. 1998).

To sum up the results of the literature review, one can easily observe that undergraduate GPA, MCAT scores, gender, and race are frequently used as significant predictors for the USMLE Step 1 performances in medical schools.

Methodology

Since 1994 the College has used a computer-based student tracking system for tracking medical student progression during matriculation and beyond graduation. This system captures individual student profiles in pre-admission variables, course performances, licensure examination results, post-graduate training, and alumni physician practicing specialties. Because of the availability and accessibility of the tracking system, institutional researchers are able to merge relevant files and retrieve the needed data to build the prediction models successfully.

The sample (N=216) was confined to the four-year (1992 to 1995) matriculants who had taken the USMLE Step 1 June examination for the first-time from 1994 to 1997. These students were categorized as 49% (105/216) males and 51% (111/216) females; 82% (178/216) African Americans and 18% (38/216) other ethnic group; 52% (112/216) HBCU graduates and 48% (104/216) Non-HBCU graduates; 81% (174/216) four-year track and 19% (42/216) five-year track.

The dependent variables in the study were the USMLE Step 1 June first-time taker pass/fail status and test scores depending on either logistic or linear regression methods. Fifteen variables were treated as independent variables—age, gender (0-male; 1-female), ethnicity (0-African American; 1-Non-African American), HBCU status (0-HBCU graduate; 1-Non-HBCU graduate), curriculum track (0-four years track; 1-five years track), undergraduate BSA, undergraduate GPA, MCAT scores (biological sciences, physical sciences, and verbal reasoning), medical school freshman GPA, numbers of basic sciences courses failed in the second year of curriculum, and

financial aid scholarship/grant, work-study, and loan amounts. These variables were selected from the medical student tracking system because they were quantifiable predictors.

Both logistic and linear regression techniques were applied to build the prediction models. These two techniques allowed institutional researchers to identify significant predictors, estimate the magnitude effects of these predictors, and perform predictions for the prospective test takers.

In logistic regression, a stepwise method was used to select the important variables having the largest Wald statistic in each step (default p value = .05 for variable inclusion). If the added variable did not significantly contribute to the prediction of passing the USMLE Step 1, then the variable would be excluded from the equation in the subsequent step (default p value = .10 for variable removal). The iteration process for selecting variables was completed when no additional variables met entry and removal criteria.

In linear regression, a stepwise method was also used to select the independent variables having the largest partial correlation in each step (default p value = .05 for variable inclusion). The general principle of variable selection and removal in linear regression is similar to that of logistic regression. If the added variable did not significantly contribute to the prediction of the USMLE Step 1 score, then the variable would be removed from the equation in the subsequent step (default p value = .10 for variable removal). The process of variable selection was completed when no additional variables met entry and removal criteria.

Study Results

(1) Logistic Regression Analysis

Using an estimated probability value of .5 as a cutoff point, the prediction accuracy for the pass group was 89%; the prediction accuracy for the fail group was 63%; and the overall prediction accuracy for the combined pass and fail group was 80%. All logistic regression coefficients in the final equation were significantly different from zero at the .05, .01, or .001 significance levels using the Wald tests; and more importantly, the p-value (0.9393) of minus two times the log likelihood (-2LL) test and the p-value (0.6939) of the model goodness of fit test indicated that the model fitted data very well. Logistic regression method yielded the prediction model for the success or failure of USMLE Step 1 June first-time takers:

Probability (Passing USMLE Step 1) = Exp (Z) / (1 + Exp (Z)),

where Exp is the base of the natural logarithm and

$$Z = -8.8838 + 0.4162 \times \text{MCAT biological sciences score} + 0.3412 \times \text{MCAT physical sciences score} - 2.1815 \times \text{number of sophomore courses failed} + 1.8672 \times \text{medical school freshman GPA}$$

The study results disclosed considerable information concerning relationships among variables in the model (see Table 1). MCAT biological sciences score, MCAT physical sciences

score, numbers of sophomore courses failed in the basic sciences curriculum, and medical school freshman GPA were used to predict the USMLE Step 1 success or failure status in the model.

Of the fifteen independent variables used in the model, age, gender, ethnicity, HBCU status, curriculum track (4- or 5-year track), undergraduate BSA, GPA, MCAT verbal reasoning score, and financial aid scholarship/grant, work-study, and loan amounts were excluded because these variables were already determined by the Wald test not to be useful in predicting the USMLE Step 1 pass or fail status.

Table 1

Logistic Regression Analysis for Predicting USMLE Step 1 Pass/Fail Status

Variables in the Equation	Logistic Regression Coefficients (B)	p Values	Odds or Exp(B)****
MCAT biological sciences score	0.4162	0.0022**	1.5162
MCAT physical sciences score	0.3412	0.0347*	1.4066
Number of sophomore courses failed	-2.1815	0.0125*	0.1129
Medical school freshman GPA	1.8672	0.0001***	6.4704

Constant	-8.8838	0.0001***

* Regression coefficient is significantly different from zero at the 0.05 significance level using the Wald test.

** Regression coefficient is significantly different from zero at the 0.01 significance level using the Wald test.

*** Regression coefficient is significantly different from zero at the 0.001 significance level using the Wald test

****the base of the natural logarithm for regression coefficient (B)

In logistic regression analysis, the logistic regression coefficients in (B) column as shown in Table 1 were interpreted as change in the log odds of passing licensure examination for every unit of change in the predictors when holding other variables as constants. If B coefficients were positive, then odds were greater than 1, indicating that the odds of passing the USMLE Step 1 increased. On the contrary, if B coefficients were negative, then odds were less than 1 and greater than zero, suggesting that the odds of passing the USMLE Step 1 decreased. Again, if B coefficients were zero, then odds became one, showing the odds of passing the USMLE Step 1 was not better than the chance of getting a head or tail when tossing a fair coin.

The interpretation of logistic regression analysis involved two parts: (1) determining the functional relationship between the significant explanatory variables and the probability of passing Step 1, and (2) defining the units of change for the explanatory variables affecting on the probability

of passing Step 1. In this study, the four significant predictors affected the magnitudes of change in the log odds (logit) and odds of passing the USMLE Step 1 when holding other predictors as constant. For instance, when the medical school freshman GPA, MCAT biological sciences score, or MCAT physical sciences score increased 1 point, the log odds of passing USMLE Step 1 were increased by a 1.87, 0.42, and 0.34, respectively, as shown in the (B) column. Also, the log odds of passing USMLE Step 1 decreased by a 2.18 when the number of sophomore courses failed increased by one. As another example, when the medical school freshman GPA, MCAT biological sciences score, or MCAT physical sciences score increased 1 point, the odds of passing USMLE Step 1 were increased by a factor of 6.47, 1.52, and 1.41, respectively, as shown in the Exp (B) column. The odds of passing USMLE Step 1 increased by a factor of .11 when the number of sophomore courses failed increased by one. The logistic regression equation in this study met the overall standard criteria for being a reasonable and workable model except the prediction accuracy (63%) for the fail group not being reasonably high (see Table 2).

Table 2

Checklists for Being a Reasonable and Workable Logistic Regression Model

Criteria	Check
Does the model fit the data well? (Is the p value for –2 times log likelihood or chi-square test statistic greater than the .05 significance level?)	Yes; Yes

Are logistic regression coefficients significantly different from zero using the Wald test?	Yes
Are signs (+ or -) of logistic regression coefficients appropriate?	Yes
Do the magnitude effects (odds) of explanatory variables make sense?	Yes
Is the prediction accuracy for combined pass and fail group reasonably high?	Yes
Is the prediction accuracy for the pass group reasonably high?	Yes
Is the prediction accuracy for the fail group reasonably high? (*As a default cutoff point, the estimated probability value .5 needed to be adjusted to either .4, .6 or other values to improve the prediction accuracy)	No*
Are residuals normally distributed with mean zero (When n is large, the normal distribution is a reasonable approximation to the binominal distribution)	Yes
Does the standard deviation of the residuals equal to one?	Yes
Are there any correlation among independent X variables? (or collinearity?)	No

---

(2) Linear Regression Analysis

In linear regression analysis, the overall prediction accuracy for the combined pass and fail group was 79%. The linear regression model appeared to be a good fit because the coefficient of determination, R square, was reasonably high (.67); the standard error of the prediction was fairly small (15 points); the population regression coefficients of variables in the regression equation were significantly different from zero at the .01 or .001 significance levels using t tests; and perhaps,

most importantly, the underlying assumptions of least-square regression such as linearity, normality, and independence were not violated. For example, variance inflation factors and collinearity diagnostic checks revealed no collinearity among independent variables; the histogram and scattergram of standardized residuals exhibited the normal curve with mean zero and constant variance; scatterplots of standardized residuals against the predictive scores and other independent variables displayed random patterns, indicating the existence of independence. In addition, the Durbin-Watson test indicated that there was no series correlation among residuals. Linear regression method yielded the prediction model for the scores of USMLE Step 1 June first-time takers as follows:

Estimated USMLE Step 1 score = 84.262 + 3.192 x MCAT biological sciences score
+ 1.934 x MCAT physical sciences score
+ 1.841 x MCAT verbal reasoning score
- 9.972 x number of sophomore courses failed
+ 0.009 x work-study dollar
+ 20.833 x medical school freshman GPA

The study revealed information about the relationship between the Step 1 performance and predictors under the investigation (see Table 3). Of the fifteen independent variables used in the model, nine variables--age, gender, ethnicity, HBCU status, curriculum track (4- or 5-year track), undergraduate BSA and GPA, scholarship/grant amount, and loan amount had no significant contribution to the USMLE Step 1 performance. The three MCAT scores, the number of basic

sciences courses failed in the second-year curriculum, work-study dollar, and the medical school freshman GPA were used to predict USMLE Step 1 scores.

Table 3

Linear Regression Analysis for Predicting USMLE Step 1 Scores

Variables in the Equation	Linear Regression Coefficients (B)	Standardized Regression Coefficient (Beta)	p Values
MCAT biological sciences score	3.192	0.270	0.0001***
MCAT physical sciences score	1.934	0.153	0.0110*
MCAT verbal reasoning score	1.841	0.143	0.0016**
Number of sophomore courses failed	- 9.972	-0.169	0.0001***
Work-study dollar	0.009	0.086	0.0339*
Medical school freshman GPA	20.833	0.431	0.0001***
Constant	84.262		0.0001***

* Regression coefficient is significantly different from zero at the 0.05 significance

level using the t test.

** Regression coefficient is significantly different from zero at the 0.01 significance level using the t test.

*** Regression coefficient is significantly different from zero at the 0.001 significance level using the t test.

The standardized regression coefficients were compared to make judgment about the relative influence of independent variables in the regression model. As shown in Table 3, the medical school freshman GPA had a standardized regression coefficient of .431 which was approximately twice those absolute values of MCAT biological sciences (.270) and number of sophomore courses failed (-.169); and nearly three times more than those scores of MCAT physical sciences (.153) and verbal reasoning (.143). When holding other predictors as constant, the magnitude changes in the USMLE Step 1 performance affected by specific predictors are listed as follows: (a) an increase of 1 point in the MCAT biological sciences score was directly associated with an increase of about 3 points in the USMLE Step 1 score, (b) an increase of 1 point on the MCAT physical sciences or verbal reasoning score is directly related to an increase of nearly 2 points in the USMLE Step 1 score, (c) an increase of 1 course failed in the sophomore year was directly tied to a decrease of almost 10 points in the USMLE Step 1 score, (d) an increase of one thousand dollars in work-study amount affected an increase of 9 points in the USMLE Step 1 score, and (e) an additional gain of 1 point in the medical school freshman

GPA predicted an increase of nearly 21 points in the USMLE Step 1 score. It was expected that the MCAT scores, medical school freshman GPA, and work-study dollar were positively correlated with the USMLE Step 1 score; and that the number of second-year courses failed was negatively correlated with the Step 1 score.

The model structure of linear regression was almost identical to that of the logistic regression except the linear regression model retained two additional variables--MCAT verbal reasoning score and work-study dollar. In the linear regression model, the regression coefficients of three MCAT scores, medical school freshman GPA, work-study dollar, and courses failed in the sophomore year were significantly different from zero. Using a standard passing score of 176 as cutoff point, the prediction accuracy for the pass group was 80% that was 9% less accurate than the logistic regression (89%). The prediction accuracy for the fail group was 73% which was 10% more accurate than the logistic regression (63%). The overall prediction accuracy (80%) of the linear regression for the combined pass and fail group was identical to that of the logistic regression model. Therefore, the linear regression model was considered as the better model generated in this study to predict the USMLE Step 1 performance. The linear regression equation in this study met the overall standard criteria for being a reasonable and workable model (see Table 4)

Table 4

Checklists for Being a Reasonable and Workable Linear Regression Model

Criteria	Check
Is the R square (measure of success of the linear regression equation in explaining the variation in data set) reasonably high?	Yes
Is the standard error of the estimation reasonably small?	Yes
Are regression coefficients significantly different from zero using F or t tests?	Yes
Are signs (+ or -) of regression coefficients appropriate?	Yes
Do the magnitude effects (slopes) of significant predictors make sense?	Yes
Is the overall prediction accuracy for the combined pass and fail group reasonably high?	Yes
Is the prediction accuracy for the pass group reasonably high?	Yes
Is the prediction accuracy for the fail group reasonably high?	Yes
Are residuals normally distributed with mean zero by checking histogram?	Yes
Do residuals display the constant variance pattern by checking scattergram?	Yes
Does the casewise residuals plot exhibit a random (independence) pattern?	Yes

Are there any outliers on the casewise-standardized residual plot?	No
Is there any series correlation or dependency by using the Durbin-Watson test?	No
Are the estimated Y scores correlated with residuals? (or systematic error?)	No
Are there any correlation among independent X variables? (or collinearity?)	No

Implications and Limitation

The medical school freshman GPA and sophomore course performance in the basic-sciences curriculum were unique predictors. MCAT scores were significant predictors of the medical licensure examination, regardless of scholarship/grant and loan amounts; medical school curriculum track (4- or 5-year); and other pre-admission variables such as age, gender ethnicity, HBCU status, and undergraduate BSA and GPA scores. This finding seemed to be consistent with the 1996 Swanson's study indicating that MCAT scores alone provided more accurate predictions of the USMLE Step 1 performance. In this study, the standardized regression weight (or odds) of the MCAT biological sciences score is higher than that of other MCAT scores. Also, on average, each one-point increase in three MCAT scores resulted in a total of 6.967 points (sum of three MCAT slopes) increase in the USMLE Step 1 score. These findings also seemed to be consistent with the 1998 Swanson's study. It is obvious that medical school freshman GPA and number of the second-year courses failed were strongly correlated with the USMLE Step 1

scores in positive and negative directions, respectively, suggesting that basic science disciplines had the predictive power for the medical licensure examination.

In comparison to the R squares ranging from .26 to .52 in the published articles, this study contained a high value of R square .67, indicating that the model fitted the data quite well and exhibited the explanatory power. The linear regression model demonstrated a higher predictive validity and produced a high degree of overall prediction accuracy (80%) in predicting the pass and fail status of the USMLE Step 1 performance. In addition, there were more independent variables involved in the model construction in this study as compared to its counterparts of the prediction models in the literature. Furthermore, this study applied logistic and linear regression methods to examine the consistency or reliability of model structure; determine the predictive validity or accuracy of the prediction models; and estimate the magnitude effects (slopes and odds) of the independent variables. More importantly, a great deal of effort was involved in checking the model assumptions and assuring the model goodness of fit. However, to achieve the highest degree of prediction validity for licensure examination performance, institutional researcher would require more quantifiable variables such as student motivation, faculty effort, college learning environment, and parents' income and education levels. These variables were not available for individual students at the time of conducting this study.

The knowledge gained from this study would be beneficial to a medical school and its students. The medical school could determine its program effectiveness based on the significant predictors found--medical school freshman GPA, sophomore course performance, and financial aid work-study dollar. The administrators of the medical school and the staff of academic support programs could use the prediction results to identify a group of potential 'at-risk' students to implement their mandatory board-review or tutorial programs. In addition, the prediction results could help the College build a consensus that MCAT scores were significant predictors of USMLE Step 1 performance and the College should continue its efforts in admitting medical students with high MCAT scores. Clearly, the admissions committee could screen student applicants for interview based on the rank order of standardized regression coefficients or odds of passing licensure examination. The USMLE Step 1 scores could be improved if some students used the prediction results to determine the optimum time to take the licensure examination. The prediction models could help the college determine and document its effective basic sciences curriculum and increase the likelihood of student success in medical education.

References

Draper, N.R and Smith, H. (1981). Applied Regression Analysis. New York: John Wiley & Sons, Inc.

Elam C. (1994). NBME Part I versus USMLE Step 1: Predicting Scores Based on Preadmission and Medical School Performances. Journal of Academic Medicine, Vol. 69.

Erdmann, J. B. (1993). Guidance for the Use of the USMLE in Medical Education Settings. Journal of Academic Medicine, Vol 68, 732-733.

Fadem, B., Schuchman, M., and Simring, Steve S. (1995). The Relationship between Parental Income and Academic Performance of Medical Students. Journal of Academic Medicine, Vol 70, No. 12.

Hosemer, David W., and Lemeshow, Stanley, (1989). Applied Logistic Regression. New York: John Wiley & Sons, Inc.

Koenig, J.A. and Wiley, A., (1996). The Validity of the Medical College Admission Test for Predicting Performance in the First Two Years of Medical School. Journal of Academic Medicine, Vol. 71; S83-S85.

O'Donnell, M.J., et al. (1993). Background Essential to the Proper Use of Results of Step 1 and Step 2 of the USMLE. Journal of Academic Medicine, Vol. 68.

Shen, S. and Comrey, A.L. (1997). Predicting Medical Students' Academic Performances by Their Cognitive Abilities and Personality Characteristics. Journal of Academic Medicine, Vol. 72; 781-786.

Silver B. and Hodgson, C. S. (1997). Evaluating GPAs and MCAT Scores as Predictors of NBME-I and Clerkship Performances Based on Students' Data from One Undergraduate Institution. Journal of Academic Medicine, Vol. 72; 394-396.

Swanson, David B. and Ripkey, Douglas R., Case, Susan M., and the 1994-95 Validity Study Group for USMLE Step 1/2 Pass/Fail Standards (1996). Preliminary Study of the Accurate of the Old and New Medical College Admission Tests for Predicting Performance on USMLE Step 1. Journal of Academic Medicine, Vol. 71, No. 1.

Swanson, David B., et al. (1998). Relationship between Achievement in Basic Science Coursework and Performance on 1994 USMLE Step 1 Test Administrations. Journal of Academic Medicine, Vol. 71.